

School of Computing and Information Systems  
The University of Melbourne  
COMP90049 Knowledge Technologies (Semester 1, 2019)  
Workshop exercises: Week 11

For the given dataset:

<i>ID</i>	<i>Outl</i>	<i>Temp</i>	<i>Humi</i>	<i>Wind</i>	<i>PLAY</i>
TRAINING INSTANCES					
A	s	h	h	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	c	n	T	?
H	s	m	h	F	?

- Find the entropy (the distribution of attribute values) for each of the six attributes.
- Classify the test instances using a Decision Tree along with each of the following approach:
  - Using Information Gain as a splitting criterion
  - Using Gain Ratio as a splitting criterion
  - Using Gini Index as a splitting criterion
- What is the contribution of bagging in developing a Decision Tree?
  - What is a Random Forest?
  - What advantages does a Random Forest have, with comparison to a (deterministic) Decision Tree model, or a bag of Decision Trees?